Science & Education (2024) 33:1033–1068
https://doi.org/10.1007/s11191-023-00422-x

ARTICLE

The Method of Observation in Science Education: Characteristic Dimensions from an Educational Perspective

Simon F. Kraus[1]

Accepted: 26 January 2023 / Published online: 28 February 2023
© The Author(s) 2023

Abstract
Observation is widely regarded as a fundamental way of studying nature. However, due to its complexity, proposing a comprehensive characterization is challenging. Observation is a method of studying nature that can be divided into several dimensions. Each of these dimensions can be presented as a pair of terms, which are elaborated in this paper based on the epistemological, philosophical, and educational literature. In addition, the dimensions can be assigned to two levels, which can be used to group the individual dimensions. This epistemological perspective on the method is enriched by historical examples from different periods in physics and astronomy to substantiate and illustrate the classifications. The aim of the paper is to provide an overview of the epistemological features of the method in terms of these dimensions, which may be helpful for discussions in teacher training courses.

1 Introduction

Alongside experimentation, observation is an essential method in the scientific approach (Daston & Lunbeck, 2011, p. 1). Einstein also clearly emphasized that sensory experience[1] is the starting point of physics: "By mere logical thinking, we are not able to gain any knowledge about the world of experience; all knowledge about reality starts from experience and leads into it" (Einstein, 1955, p. 113, translation by the author).

Scientific observation, as an empirical access to nature, has long been discussed from the perspectives of philosophy, epistemology, and classroom application (e.g., Heath, 1980; Kunert, 1971; Torretti, 1986). An adequate characterization of observation for school purposes (both the training of teachers and the use of the term in schools) requires the integration of all these perspectives. The approach adopted here to characterize observation as

[1] Einstein's concept of sensory experience is very broad and, in contrast to a classical natural philosophical approach, also includes abstract entities, such as the constancy of the speed of light or the equivalence of heavy and inert mass (Holton, 1979). Such an extended understanding of the term avoids highly simplistic views that see observations merely as the starting point of further investigation (Hodson, 1998).

✉ Simon F. Kraus
kraus@physik.uni-siegen.de

[1] School of Science and Technology, Department of Physics, Physics Education, University of Siegen, Adolf-Reichwein-Str. 2, 57068 Siegen, Germany

comprehensively as possible focuses on two levels, taking into account the perspectives mentioned above.[2] The characterization begins with basic properties that are fundamental to the method as a scientific form of inquiry, before proceeding to the methodological-instrumental level, which opens a broad field of possibilities for more extensive methodological comparisons, such as the classification of historical investigations or the planning of reflection on lessons.

In the course of the development of science, epistemology, and philosophy, different authors in different epochs have attributed different meanings to the concept of observation. However, the present article cannot and does not intend to provide a comprehensive inventory and comparison of all these positions. Instead, it will attempt to identify those aspects that either appear significant across epochs (e.g., the personal dispositions of the observer, pairs of terms conscious/unconscious and passive/active), have a special significance for the teaching of physics in schools (e.g., observations, measurements and their relation to each other, pairs of terms qualitative/quantitative), or seem to have a high relevance for the future educational implementation in school (e.g., modes of representation in instruments, pair of terms iconic/numeric).

In order to be able to include the most diverse manifestations of empirical methods of knowledge acquisition, the definitions of these methods are usually very broad. Thus, according to Irzik and Nola (2022), observation, in its simplest form, "is the recognition and the subsequent recording of an event or fact through the five senses without the aid of any instrument." However, the authors make immediately clear that such a simple form of observation, by its very nature, promises little gain in knowledge and therefore requires the consideration of instruments. Any attempt of a definition must therefore almost inevitably remain incomplete. Since this paper will repeatedly address the relationship between observation and experiment, the definition of experiment is also relevant. Again, this must remain relatively general if the particular idiosyncrasies of each discipline are to be included. A possible definition for experiment is therefore: "[A]n experiment is an intervention into the workings of nature by using (often sophisticated) instruments, where a certain natural object or aspect of a natural phenomenon is manipulated under artificially created, highly idealized and fully controlled conditions and then its effect is observed" (Irzik & Nola, 2022). Already from these general definitions and remarks, it becomes clear that, first, observation as a method must be more complex than this simplest form and, second, both methods (the experiment or the observation) are closely interwoven. We will repeatedly address these aspects in the further course of the paper.

For the epistemological considerations made here, a general preliminary remark must be pronounced: There is a close connection between the subjects physics and astronomy and the teaching of these subjects in school. However, it must also be noted that this relationship is by no means simple and direct. Experiments and observations as an indispensable part of the process of acquiring knowledge are present on both sides. At the same time, certain educational functions exist in the school implementation of experiments and observations, for which there are not necessarily equivalents in science. Kampourakis (2016) expresses this discrepancy as follows: "However, all school science is in some way a refined, highly processed version of the actual science." Nevertheless, if one accepts that providing insights into the practices of the reference science is also one of the objectives

[2] The different perspectives of philosophy, epistemology, and classroom application, however, are not identical to the levels used here, which instead represent an attempt at a classification that proceeds from the necessary conditions for an observation to the various manifestations of this method.

of teaching in school, this is possible even without a complete correspondence between the science and the school subject. However, a systematic comparison with respect to observation cannot be made within the scope of this paper. Approaches in which the perspective of practical implementation in schools is specifically addressed for the method of observation can be found in the literature (e.g., Ford (2005); Eberbach and Crowley (2009)).

Pairs of terms are often (implicitly) used in the literature to describe and characterize observations, such as simple/complex (Norris, 1985), direct/indirect (Kunert, 1971), and systematic/unsystematic (Eberbach & Crowley, 2009). This form of representation seems appropriate as a basis for discussing each dimension, although this paper will predominantly use other pairs of terms for various reasons. However, as we will additionally show, this division should often not be understood as a dichotomous juxtaposition, as such a reduction would not be appropriate for the method.

First, however, we examine critically the evaluation of the importance of the method of observation for the sciences. This shall be done first by a contrasting example, in which just the second essential method of physics, namely the experiment, is emphasized with respect to its special role in the acquisition of knowledge. Already, Otto von Guericke (1602–1686) argued that "more weight is to be put on experiments than on the judgment of stupidity, which always tends to spin prejudices against nature" (as cited in Slaby, 1907, p. 21 translation by the author). Experiments appear here as the "gold standard" in science. Why, then, should observation, as another method, be given any attention from an educational point of view? Von Guericke's quote can also be understood as a general call to turn to an empirical approach, which must be understood in the context of the development of modern sciences as von Guericke was a contemporary of essential methodological innovators such as William Gilbert (1544–1603) or Galileo Galilei (1564–1641). The statement does not contradict an observation-based approach, which also claims not to "spin prejudices against nature" but to investigate it objectively. How this can be achieved has been demonstrated for millennia in the field of astronomy and for over a century in astrophysics. As we will explain in the following, experiments themselves are not conceivable without observations; consequently, a comparison of the two methods in this simple way does not appear meaningful. Both methods are indeed sources of empirical information.

Von Guericke's quote expresses a likely common notion in disciplinary education and teaching regarding the methodological focus of investigations in science. However, this notion (Furtak & Penuel, 2019; Hodson, 1998; Ioannidou & Erduran, 2021), referred to as the "primacy of experimentation" (Emden, 2021), creates a bias in the evaluation of each approach to nature. The objective of this paper is therefore to elaborate a view of observation that can be considered viable and that constitutes a suitable basis for communicating the characteristics of the method, for instance, to student teachers. In particular, the tension between the different perspectives on observation should be made clear to reveal the complexity of this method of knowledge acquisition and transfer it, at least partially, into the classroom. The importance of developing appropriate conceptions of nature of science for practical teaching is also highlighted in the literature (e.g., Haslam & Gunstone, 1998; Ioannidou & Erduran, 2021).

2 Levels of Characterization

Norris (1985) points out three basic misconceptions about observation as a method in the science education literature:

- The existence of a general difference between observation and interpretation
- The close connection of observation to the human senses
- The assumption that observation is one of the least intellectually demanding methods

As Norris emphasizes, it can be demonstrated that none of the three assumptions is tenable under the conditions of modern science. As we will show in the course, these assumptions belong to different levels and are not sufficiently differentiated to provide a precise characterization of the method. In the following, we present a first level on which the characterization of observations can take place. In the further course, the various dimensions that can be assigned to this first level will then be elaborated. Among other things, it will be shown which examples could be used to counteract the misconception pointed out by Norris.

2.1 The Fundamental Level

The fundamental dimensions are those that enable the characterization of observation as a scientific method and serve to distinguish it from the everyday use of the term (Eberbach & Crowley, 2009). The four pairs of terms listed here can only be occupied one-sidedly in each case; otherwise, it would not be a scientific[3] observation under certain circumstances. The characterization of observation as a method of knowledge acquisition based on these or similar properties is undisputed in the literature and is reproduced in the interest of completeness.

2.1.1 Unconscious/Conscious

One of the most elementary statements about scientific observations is that seeing alone does not constitute an observation (Eberbach & Crowley, 2009). As an example, Galileo Galilei observed Jupiter and its moons on January 4, 1613. In the graphical records of this observation, besides Jupiter and all four Galilean moons, another entry appears in the form of a star symbol. At the time of the observation, the still-unknown planet Neptune was at the precise location where Galileo registered a fixed star in his star chart. However, there is no indication that Galileo saw anything other than a star in this object, which later[4] turned out to be the eighth planet of the solar system. Neptune was within an angular distance of Jupiter much smaller than the maximum angular distance of the four large moons (Fig. 1). A single glance through the telescope could thus have given the impression that the apparent fixed star was part of the Jupiter system. Since Galileo based his investigations on repeated observations, and especially on the detailed study of the variable orientation of the moons in relation to Jupiter on different nights, he paid no further attention to the object. Because of its immense distance from the Sun, it had no visible extent in Galileo's small telescope, and it did not shift immediately with respect to the stellar background due to its long orbital period. Implicitly, he thus classified the object as an ordinary star (Hermann, 2004; Kowal & Drake, 1980).

[3] In most cases, even a colloquial notion of observation is likely to be inapplicable when, for example, a fact is not perceived consciously.

[4] Neptune was not discovered until 1846, by Johann Gottfried Galle.

Fig. 1 Illustration of the principle alignment of the celestial bodies as it appeared at the time of Galileo's observation of Jupiter. The moons can already easily be identified as being in motion over short periods

This example shows that the mere transmission of information from a sender to a receiver—here, in the form of light to the human eye—is by no means sufficient to speak of an observation. Only the classification of an object or phenomenon as something remarkable, which goes beyond the pure visual process, transforms what is seen into a conscious perception. The transition is thus inextricably linked to situational interest (Hidi, 2001; Knogler, 2017), which must be aroused to prevent the information from being lost in the background noise of sensory impressions. Thus, observation is controlled by knowledge (possibly only applied intuitively) as well as by the expectations and previous experiences of the observer.

2.1.2 Passive/Active

Indisputably, an observation requires some form of activity on the part of the observer (e.g., Abelmann, 1965; Heath, 1980; Kim & Park, 2018; Norris, 1985). However, as we will argue later, the notion of activity cannot refer to a purely physical form, which would already cause a change in the conditions of observation (e.g., a shift of the head to look at a flower from a different angle). It is first of all a purely mental activity.

Distinguishing between observation as a procedure of empirical knowledge acquisition and the more general concept of a "sensory-rational process" (Kunert, 1971, p. 180), Grimmer resorts to attention as an essential element of the former:

> Observing means... watching and mentally grasping an ongoing process, whereby the observer's attention is not necessarily directed to the overall appearance, but rather – in a planned and problem-conscious manner – to the essentials of the appearance at the respective moment and the circumstances associated with it. (Grimmer, as cited in Kunert, 1971, p. 181, translation by the author)

The keywords "planful" and "problem-conscious" also refer to an activity that, in many cases, must precede the procedure, especially if the phenomenon proceeds rapidly. This pair of terms is closely related to another pair, exploratory/validating, through which the preconditions for the activity in the practical implementation in the classroom are created. However, the decisive factor at this point is the attention of the observer, which implies a mental activity and does not already aim at a methodical decision.

The emergence of modern empirical sciences was also accompanied by several changes in the emotional attitude toward the object of investigation. While Aristotle saw the beginning of philosophy in amazement, medieval natural philosophy turned away from this view and perceived amazement as an expression of ignorance that ought to be avoided (Daston, 2001). In contrast, intensive attention to the phenomenon, which should also include the emotional dimension, again represented the basis of knowledge for Newton, for example. Daston summarizes Newton's sequence as follows, from the perception of the phenomenon to its communication to the public:

> Newton, in his 1672 report on the beginnings of his exploration of spectral colors, adhered to a precise sequence: first there was wonder, then came attention, and finally curiosity. (2001, p. 25, translation by the author)

Robert Hooke put great emphasis on drawing attention to everyday objects, which he explored with the novel microscopes by presenting them as amazing:

> In all kinds of observations one should exercise a great deal of circumspection so as not to miss even the smallest perceptible circumstance… And an observer should

> undertake to look at experiments and observations which are more ordinary, and with which he is more familiar, as if they were the greatest rarity, and imagine himself to be someone from another country or other profession who had never seen or heard anything similar before. (Hooke, as cited in Daston, 2001, pp. 25-26, translation by the author)

Studies of the observational ability of learners, meanwhile, show that the latter arbitrarily take note of partial phenomena and processes. Since attention does not focus on the relevant elements even with increasing age, the notion that this effect is a simple consequence of the level of cognitive development can be rejected (Eberbach & Crowley, 2009).[5] The observed tendency to go no further than superficial features is instead attributed to a lack of subject-specific knowledge (Chi et al., 1989; Eberbach & Crowley, 2009; Johnson & Mervis, 1994, 1997).

Actively directing attention to the relevant details of a phenomenon thus requires more than what is achieved by evoking a sense of wonder. The necessity of going beyond the arousal of curiosity and situational interest also results from the fact that observation occasions that are limited to short-term attention and neglect the simultaneous activation of prior knowledge only contribute to the acquisition of new knowledge to a very limited extent (Eberbach & Crowley, 2009; Tomkins & Tunnicliffe, 2001):

> When children are cast into an activity with inadequate knowledge and instructional support, observation becomes a weak method for collecting data rather than a powerful method for reasoning scientifically. (Eberbach & Crowley, 2009, p. 49)

Thomas Kuhn also refers to the close link to prior knowledge when he states that an observation requires ascertaining "both that something is and what it is" (1962, p. 762).

However, since students are not likely to possess in advance the necessary background knowledge for many tasks, it is necessary to direct their attention externally to the essential elements and phenomena. In demonstration experiments, such guidance by appropriate design and by highlighting central components is part of the usual procedure in class. It can also be supplemented by active guidance through appropriate instructions or carried out via observation tasks besides experiments. A free approach termed "exploratory learning"[6] is to be seen critically since high learning effectiveness cannot be expected from such approaches, for the reasons mentioned above (Kirschner et al., 2006).

2.1.3 Singular/Repeated

The historical example of Galileo presented above shows that even a conscious perception of a sensory stimulus (here, in the form of light from the supposed star) does not necessarily have to lead to a discovery. In Galileo's case, observing Jupiter and its moons repeatedly was decisive to distinguish the moons from faint stars as this was simply not possible

[5] In the studies cited, instead of observation itself, we consider the interpretation of the observational results or the ability to generate coherent and connected explanations based on the observations. If we continue to follow Norris's (1985) view and do not introduce a sharp distinction between observation and interpretation, this fact does not contradict our argument in favor of a view of observation as a process that requires activity.

[6] See also Section 2.2.1, which considers the similar but detailed and independent exploratory approach in science.

for Neptune. However, the necessity of repeating observations is made even clearer by the example of Wilhelm Conrad Roentgen's discovery of a new kind of radiation.[7]

Roentgen's experiments with cathode rays caused one of his screens coated with barium platinocyanide to fluoresce even though it was shielded from the direct cathode radiation of the tube. Roentgen took this as an opportunity to investigate the novel phenomenon in more detail and quickly came to the conclusion that his discovery must reveal a new type of radiation: the X-rays later named in his honor. Reports confirm that he was able to observe the interaction of the new radiation with the screen several times and then immediately subjected the former to various kinds of experiments in rapid succession. What it means, however, if a phenomenon is observed only once, can only be clarified if one takes a contrasting example.

The great difficulty here is that an accompanying "non-discovery" is not as well documented and investigated as a discovery. In the case of X-rays, however, Herbert Jackson, Johann W. Hittorf, Eugen Goldstein, and Philipp Lenard saw luminous phenomena during their experiments with cathode rays. However, these luminous phenomena on various materials did not receive further attention from these scientists because they each tried to exclude observations that appeared irrelevant to their specific core questions (Glasser et al., 1933, p. 224). After all, Lenard stated that he was quite aware of the phenomenon and even noted it for later investigations: "In fact, I had made several inexplicable observations, which I carefully preserved for later investigations, unfortunately not begun in time, and which must have been the effects of traces of wave radiation" (cited in Glasser et al., 1933, p. 224, translation by the author).

Thus, while Lenard and others each devoted their attention to a particular question to which the fluorescence of the screen was irrelevant, we can assume that Roentgen benefitted from flexibility (see Emden, 2021) that allowed him to notice the specificity of the screen's reaction to radiation from the tube radiation even given the general novelty of the technology of cathode ray tubes. Moreover, he was ready to abandon his original activity and immediately turn to the new phenomenon to study it in detail (Simonyi, 2012, p. 380). Thomson made an observation in comparable conditions:

> I have been able to detect phosphorescence in pieces of ordinary German-glass tubing held at a distance of some feet from the discharge-tube, though in this case the light had to pass through the glass walls of the vacuum-tube and a considerable thickness of air before falling on the phosphorescent body. (Thomson, 1894, p. 359)

Repeated observation implies that the process under study can be triggered actively and purposefully by the observer and corresponds to the concept of reproducibility as one of the characteristic properties of the experiment. In the course, we will show when and under which conditions a classical reproduction[8] is possible by further breaking down the concept of observation. At this point, it should suffice to state that observation can be understood as an active act, as described above. Thus, the active search for a similar object or phenomenon or the willingness to place the phenomenon in the context

[7] With regard to the chosen example of Roentgen's discovery, the objection could be raised that this is a phenomenological experiment and not or not exclusively an observation. As will be shown in the further course (see Section 2.2.2), an observation can also occur within an experiment without losing its character as an independent method.

[8] "Reproduction" is used here in the sense of an actively controlled repetition, as it is performed in the experiment.

of previous investigations when it occurs again by chance can also be seen as a form of reproducibility.

What happens when such novel results cannot be repeated is illustrated by the same historical example? Here, we focus on Roentgen's precariousness when he addressed the question of the direct perceptibility of the new radiation by the naked eye (see Glasser et al., 1933):

> The fact that Mr. G. Brandes observed that the X-rays can trigger a light stimulus in the retina of the eye, I have found confirmed. Also in my observation journal there is a note from the beginning of November 1895, according to which in a completely darkened room close to a wooden door, on which exterior a Hittorf's tube was attached, I perceived a faint light phenomenon extending over the entire field of vision, when discharges were sent through the tube. Since I observed this phenomenon only once, I considered it subjective, and the fact that I did not see it repeatedly is due to the fact that later, instead of [a] Hittorf's tube, other less evacuated, non-platinum anode devices were used. (Röntgen, 1897, pp. 591–592, translation by the author)

As it becomes clear in retrospect, Roentgen could not repeatedly perceive the radiation because of the insufficient intensity produced by the tubes that he used. In contrast to the situation of the original discovery of X-rays and his otherwise extensive studies of all properties of radiation, he paid no attention to this phenomenon. It can hardly be assumed that the preceding research on radiation led to a change in Roentgen's disposition, which was described above as characterized by flexibility. Rather, the peculiarity of the method of observation can be revealed here. We will return to the different forms of observations in the course (see Section 2.2.2).

Even a repeated observation using the same simple and repetitive procedure should not be given excessive confidence if the detection is at the limit of the (if necessary, instrument-supported) perception. Such confidence led to the supposed discovery of exoplanets around Barnard's star by Peter van de Kamp (Harwit, 2021, p. 199). Over 43 years, van de Kamp measured the star's position in the sky approximately 1200 times and deduced a slightly periodic displacement of the star's orbit in its proper motion. From these measurements, he concluded the existence of two planets with masses of 0.7 and 0.5 Jupiter masses, respectively, and orbital periods of 12 and 20 years, respectively. However, this work approached the limits of the telescope's resolving power and overemphasized random and systematic errors of the instrument while a substantial technical advancement failed to materialize. Today, the existence of these planets is rejected as a misinterpretation.

In addition to repeatability as an essential criterion of quality for observational data, this section highlighted the importance of a certain degree of mental flexibility to allow accidental discoveries. Here, the personality of the researcher is a crucial factor (Emden, 2021; Grinnell, 2013; Selby, 2006). Recognizing a lucky coincidence as such cannot be learned (Brown & Kumar, 2013; Emden, 2021) but, at the same time, is an essential insight into the process of knowledge acquisition in science (Emden, 2021).

The scientist's prior knowledge is another necessary condition for discoveries. This is affirmed, for example, by Louis Pasteur: "In the field of observation fortune favors only the prepared mind" (cited in Emden, 2021; Gillies, 2006). Neither comprehensive knowledge of the field nor flexibility based on it and closely linked to methodical approaches can be presupposed or directly learned by students. Science education must therefore initially limit itself to presenting the approaches to research subjects to develop an accurate view of

the nature of science. In doing so, the importance of personalities should be appreciated, and the difference between flexibility and arbitrariness should be emphasized.

2.1.4 Unbiased/Biased

Observational biases can be assigned to two different categories. First, biases can be caused individually by the perceptual apparatus. Second, biases that are independent of the perceptual apparatus can be caused by the observer's beliefs.

An example of the first category of distortions is the personal equation, where the measurement of the transit times of stars through the center of the eyepiece is subject to a constant individual deviation from the actual transit time (Schaffer, 1988). Another distorting effect, which unfolds most essentially in the field of astronomy, is the Purkinje effect. Here, an individual misjudgment of star brightness occurs depending on the color of the comparison star (Duncombe, 1945).[9] For both examples, the bias cannot be determined unless suitable technical equipment is available that allows a direct comparative measurement.[10] Both types of biases have in common their invariance in time and lack of sensitivity to varying external influences. Therefore, the presence of a bias is only determined by direct comparison with other observers. Consequently, reproducibility alone cannot be sufficient as a criterion for assessing the quality of an observation and its meaningfulness.

The second category of biases includes theoretical beliefs, which can be understood, in line with Thomas Kuhn (1962), as paradigms or related concepts. Such paradigms are not merely effective on the level of interpretation but rather shape perception itself. The early investigations of electrostatics can be used as an example of this (Kuhn, 1962, p. 159f.). William Gilbert, among others, studied magnetism and electrostatics and was able to show that electric charges cause attraction. However, he was not aware that repulsion could also be caused by electrical forces (Simonyi, 2012, p. 321). Although these phenomena became more clearly recognizable only through technical advancements, which were achieved primarily by Hauksbee, repulsions were in principle observable before. Nevertheless, they were seen as a purely mechanical rebound unrelated to the electrostatic phenomena. As the prevailing paradigm did not envisage repulsion resulting from electrical forces, this kind of repulsion was not detected (Kraus, 2018; Kuhn, 1962). Kuhn emphasizes that shift in the viewpoint is not a change in the interpretation of observations but a change in what is seen:

> Rather than being an interpreter, the scientist who embraces a new paradigm is like the man wearing inverting lenses. Confronting the same constellation of objects as before and knowing that he does so, he nevertheless finds them transformed through and through in many of their details. (1962, p. 122)

Since such directing of attention and perception along paradigms is constant, no observation can exist that is not theory-driven. Thus, observation must always be regarded as

[9] In a comparison of red stars with a whitish shining reference star, the red stars are comparatively underestimated in their brightness if they are faint and overestimated if their brightness is high. The effect proceeds from the different spectral sensitivity of cones and rods, which enable vision in very bright and dark environments, respectively.

[10] Purkinje was able to demonstrate on himself the transition from day to night vision and the accompanying shift in spectral sensitivity. However, the exact brightness at which this transition occurs was relevant to the astronomical brightness estimates, and this value is individually different, so that only a direct comparison can reveal and compensate for the effect.

biased. In certain circumstances, the bias of perception described at the beginning is present. Kuhn elaborates on this:

> The operations and measurements that a scientist undertakes in the laboratory are not "the given" of experience but rather "the collected with difficulty." They are not what the scientist sees—at least not before his research is well advanced and his attention focused. Rather, they are concrete indices to the content of more elementary perceptions, and as such they are selected for the close scrutiny of normal research only because they promise opportunity for the fruitful elaboration of an accepted paradigm. Far more clearly than the immediate experience from which they in part derive, operations and measurements are paradigm-determined. (1962, p. 126)

Based on this assessment, one can think back to von Guericke, who criticized the steering of perception by prejudices and pointed out the superiority of the experiment. In this sense, an observation is not guided or influenced by such prejudices or biases only if the observer is conscious of the underlying paradigm. However, a scientific investigation at the stage of normal science in Kuhn's sense does not usually exhibit such awareness (1962, p. 24). Such an objection, however, applies equally to the experiment, which is similarly subject to the influence of paradigms and, as we will show in the course, cannot do without observation.

Even theoretical presuppositions or assumptions that originate from everyday experiences can lead to a bias of perception. For example, if a student believes that the moon is only visible at night, they no longer perceive it in the daytime sky (Eberbach & Crowley, 2009; Vosniadou & Brewer, 1994). This type of perceptual bias is well known as "confirmation bias" (Klayman & Ha, 1987). Importantly, it takes place at the level of observation, not interpretation.

The use of the term pair unbiased/biased here should be examined. The term "bias" is likely to have negative connotations, both in general and in the context of a scientific observation, which uses it with reference to observation that is questionable from the perspective of the nature of science. However, the idea here is to combine both categories—prior theoretical beliefs and perceptual aspects—since both lead to bias. Because we focus on the outcome or the influence of bias on the observation, the very different causes do not matter at the moment. Clearly pointing out the biases of observation, and especially their inevitability, may be suitable to counteract the formation of an inappropriate ideal image. Heath (1980, p. 157) mentions this inappropriate ideal image, for example, when he asserts that "even if care is taken to avoid bias in observation, it is still possible that the observer's perception… of an object can be wrong." In contrast, accepting the multiple and often individual biases not only as unavoidable but also as an essential part of observation contributes to the development of a more accurate picture of the nature of the method.

2.1.5 Summary: Dimensions of Observation at the Fundamental Level

The fundamental dimensions of observation presented so far are neither dichotomous properties nor continuous progressions between two poles. Instead, a scientific observation can be unambiguously assigned to only one side of the conceptual pair. Thus, an observation must necessarily occur consciously, as Heath (1980, p. 156) presents it, starting from sensory perception: "A first step in perception is that of attending to a stimulus, and selecting some of its features and disregarding others." Closely related to this, an activity

is necessary that guides perception. The prerequisite is the existence of a consolidated "orientation knowledge," which enables the classification of what is observed into existing knowledge structures and, thus, allows for a purposeful directing of attention. Contrary to earlier ideas (Heath, 1980; Wachtel, 1967), which compared focused attention to the light beam of a flashlight, recent research shows that this process is better imagined as a filter (Nakajima et al., 2019). Equally elementary—and, consequently, comparable to the requirements for a successful experiment—is the necessity of repeatability, which alone enables a confirmed observation that can be considered scientific. As the examples have shown, it can be demonstrated for scientific practice that scientists indeed proceed in this way.

A more complex concept than the fundamental dimensions addressed above is that of perceptual bias. As we have seen, contrary to the colloquial use of the term, biases are not defects of individual observations but an unavoidable part of any observation. It is not always possible to consciously perceive biases that arise from the influence of a paradigm or theory in an individual case. This cannot be an objective of teaching given that science itself does not provide such transparency or even usually pursue it. However, the existence of such influences should be addressed in educational contexts. This can be done, for instance, by using the historical examples above, emphasizing the power of paradigms so as not to evoke disdain for the achievements of earlier generations of scientists (Kraus, 2018). Biases at the level of sensory perception can also suitably intersect with the subject of biology and, thus, foster a motivating context for teaching.

2.2 The Methodological-Instrumental Level

The method of observation has so far been examined on the level of underlying perception. On this fundamental level, the method's characteristics are entirely independent of the purpose of its use (i.e., a scientific observation in the narrower sense or an observation task in an educational context). In the following, we bring those dimensions into focus on a second level referring to concrete situations, circumstances, and the instruments used (if applicable), which can generally be influenced in the design of a lesson.

2.2.1 Exploratory/Theory-Driven

The classical epistemological view of experiments assigns them the sole task of testing theoretical predictions, either validating them or (if incompatible results are produced) falsifying the theory (Popper, 1976). However, historical case studies also reveal experimental investigations whose execution does not fit this ideal of the theory-guided approach. Instead, an approach exists that does not aim to test specific expectations but rather uncover patterns and regularities for which no established theory is available. This approach is referred to as "exploratory experimentation"[11] (Burian, 1997, 2013; Steinle, 1997):

> Experiments count as exploratory when the concepts or categories in terms of which results should be understood are not obvious, the experimental methods and instruments for answering the questions are uncertain, or it is necessary first to establish

[11] Exploratory experimentation is not to be confused with "exploratory techniques" (Harwit, 2021, p. 347), which involve the on-site exploration of originally remote locations, for instance, in interplanetary space missions, and which will be briefly addressed later in this paper.

> relevant factual correlations in order to characterize the phenomena of a domain and the regularities that require (perhaps causal) explanation. (Burian, 2013, p. 720)

According to Steinle (1997), instrumental setups used in exploratory experimentation are characterized by a high degree of flexibility and allow the occurrence of unexpected results. At this point, the question arises of whether observations generally correspond to a more explorative approach due to their passive character, which does not restrict the phenomena in their appearance, and flexible use of the vast majority of (e.g., astronomical) instruments.

Table 1 presents a comparison of the central aspects of theory-guided and explorative experimentation. Based on the work of Steinle (1997), O'Malley (2007) elaborates distinctly different aspects for the two forms of experimentation.[12] We attempt to transpose these considerations to observation to determine whether two forms of observation, one theory-driven and the other exploratory, can likewise be identified. We partly use examples from astronomy, an ideal–typical observational science in which "pure observation terms" (Kuhn, 1962, p. 117) occur, which facilitate this comparison.

An example that can be seen as a form of exploratory observation strongly dominated by the instrument used is the observation of particle showers by Pierre Auger (Harwit, 2021, p. 99). The discovery of coincidences in the response of Geiger counters to photons of cosmic origin in an exploratory way led to the observation that such coincidence is preserved even when the detectors are placed further apart. Auger describes the processes in retrospect as follows:

> It was later on that I wanted to evaluate the extent of these showers by studying cosmic ray showers created in a lead screen by the observation of coincidences in two or three counters as a function of their separation. It was a surprise to observe coincidences when the counters were separated by more than one meter. Suspecting a new phenomenon, I decided to go whole hog, if I may so express myself, and thanks to the technical help of Roland Maze, we placed one of the counters in another building, more than one hundred and fifty meters away on rue Pierre Curie where my laboratory was. And there were still coincidences! It was the discovery of 'cosmic ray showers'. By persuing [*sic*!] this work at high altitudes, in order to increase the cosmic ray intensity, I showed that the showers covered more than a hectar[e] in extent, and hence the number of particles making up the showers was such that the energy of the primary particle which originated the 'giant shower' must have been more than a million billion electron-volts. That was nearly a billion times greater than the energies of particles accelerated by cyclotrons in these days. (as cited in Persson, 1996, p. 786)

[12] The complex idiosyncrasies that arise in different ways of doing science make it difficult to fully divide them into a scheme such as theory-driven and exploratory approaches without revealing contradictions in the division. For example, a setting that can be classified as exploratory may well consist of simple materials with very manageable interactions (e.g., Oersted's experiments with a conductor carrying an electric current and a magnetic needle). Conversely, theory-driven setups may be difficult to classify under the label "simple/simplified systems" (e.g., the Large Hadron Collider). Nevertheless, the grid seems to be suitable, with limitations, for describing a rough classification into certain approaches. The reference to the imperfection of the scheme in the form of the examples given and the necessity to criticize it came from an anonymous reviewer of this manuscript, who should be thanked separately for his/her comments on this issue.

Table 1 Comparison of theory-driven and exploratory experimentation (O'Malley, 2007)

	Theory-driven	Exploratory
Aim	To test specific expectations	To obtain empirical regularities; to generate concepts and classifications
Phenomena	Simple/simplified systems	Complex interacting systems
Research basis	Specific, theoretically derived research question; isolated experiment	Broad inquiry based on multiple experiments and their relationships
Theory	Well formed; guiding	Often unavailable
Instrument	Specifically designed	Flexible
Historical period	Field already formed; expanding and refining	"Epistemic situations" rather than field/tradition/period
Consequence	Theory clarification	New and fundamental conceptual framework

No existing theory predicted such interaction of cosmic particles with the terrestrial atmosphere and a cascade-like shower of secondary particles and photons in the 1930s. The extension of the distances between detectors up to 300 m and measurements on the mountains Jungfraujoch and Pic du Midi showed the extent of the cascades as well as the increase in the number of events in the thinner atmosphere in the high mountains (Auger et al., 1939). Here, a variation of certain parameters (i.e., the distance between and location of the detectors) was possible. The use of cloud chambers also allowed the first statements to be made about the nature of the secondary particles.

It is generally true for astronomy that the observation of the investigated phenomena cannot be limited to reduced systems. Since there are usually no opportunities for interaction, the system must always be studied in its full complexity. This points to a more exploratory nature of the observations and, apart from laboratory experiments on astrophysical principles (White, 2022), is characteristic of astronomy as a whole.

For O'Malley's category of scientific basis, it often follows a broad approach in astronomy described by Steinle and O'Malley as typical of an explorative approach (O'Malley, 2007; Steinle, 1997). Thus, working with archive data covering different wavelength ranges in the electromagnetic spectrum is a typical approach for exploring phenomena and interactions in detail. Observations in a specific wavelength—or simply in the visually perceptible part of the spectrum—are often part of a broader approach, in which the statistical nature of statements about astronomical objects comes into play. Nevertheless, observations can be based on theoretically narrow questions. Additionally, the use of a specific instrument does not enable a statement to be made about the presence of a particular type of observation. The combination of different wavelengths alone also does not indicate an explorative character since even theory-based observations do not provide a comprehensive insight into the nature of the phenomenon based on a small section of the spectrum due to its complexity. A classification should thus consider both the instrument and the underlying question.[13]

A broad and flexible approach also applies to the instruments in use, which are usually very flexible in their possible applications, even if one does not consider the telescope alone as the instrument but the specific combination of telescope and detector. There is only rarely a binding of the instrument to a special object as with solar telescopes. However, even in such cases, the use of the instrument is usually not restricted to a special phenomenon. In particular, the widespread archiving of observational data, which was mentioned previously, favors the subsequent use of data beyond the original purpose of the investigation. This is a peculiarity of astronomical research compared to other sciences, which may also be due to the principle limitation of the number of carriers of information, the phenomena, and the properties of the cosmos limiting the dissemination of information (see Harwit, 2021, pp. 81–196). Adding to this the longevity of astronomical objects and the slowness of processes and their development, the tracing of a single object exceeds the capabilities of very long-term archival efforts. However, newly developed instruments continue to uncover new and unexpected phenomena, such as intermediate-mass black holes, for which no comprehensive explanation is available yet. At the same time, however, the flexibility and openness of such instruments (in this case, gravitational wave detectors) do not allow for an exploration of the new phenomenon precisely because parameter variation, and in particular systematic variation, is not possible.

[13] The major role of statistics in astronomy should also be noted given that valid statements about object classes and evolutionary stages can only be made by looking at large samples.

Galison (1997) follows a comparable path in differentiating between approaches to research into a phenomenon, focusing less on the concrete actions and more on the historical classification. In Steinle's (1997) description, this classification approach would correspond most closely to the historical period, since the path to discovery only becomes visible in retrospect, once the result is available. Galison distinguishes between research results for which the research followed a "logic tradition" and that had a so-called golden event as their basis. He refers to this "golden event" as a single event yielding findings that are indisputably compelling.[14] Harwit (2021, pp. 253–279) applies this classification scheme, originally developed by Galison based on the history of high-energy physics, to astronomy.

As an example, which goes back to a "logic tradition," Harwit (2021, p. 261) cites the discovery of brown dwarfs. Decades of pondering the possible mechanisms of energy release from low-mass stars led to the realization that, based on deuterium fusion, such stars could also exist and shine for billions of years. In the 1990s, 10 Jupiter masses were calculated as the lower limit for such fusion reactions. The low surface temperature resulting from the small energy output should create a brownish color on the surface of this kind of dwarf star. The first brown dwarf was discovered by Gliese 229b in 1985, at a distance of only 5.7 pc from Earth, with a surface temperature of less than 1000 K. In the following years, the number of stars in the new spectral classes L and T increased to several hundred. The discovery can thus be seen as a pure confirmation of the theoretical predictions of the existence of brown dwarfs.

A contrasting example of a significant advance in knowledge caused by a "golden event" is the discovery of quasars (Harwit, 2021, pp. 267–268). For a long time, radio sources, such as 3C 48 or 3C273, were thought to be radio stars because of their very small angular extent. However, accurate spectral studies by Maarten Schmidt (Schmidt, 1963) revealed emission bands, which he identified as the Balmer lines of hydrogen, but these were extremely redshifted. He proposed the revolutionary explanation that the objects had to be the core regions of active galaxies, be located at great distances (in the case of 3C 273, 500 MPc was assumed), and unfold correspondingly extreme luminosities. At the same time, the core region thought to be responsible for the emissions had to be relatively small, less than 1 kPc, to explain the brightness variations observed over a few days (Hazard, et al., 2014). The discovery of an entirely new class of objects originated from this detailed examination of what was previously thought to be a known object, without any relevant theory or so much as a hint that an object like a quasar might exist.

As we have seen, in observational science, both a theory-guided approach and discovery by chance are possible. In this context, it is necessary to point out that a "golden event" is by no means a discovery "in passing." Discovery is also the result of a systematic and elaborate campaign, but it is not preceded by any theoretical preparatory work fundamental for the result.

Not every "golden event" has to be directly linked to the introduction of a new instrument (Harwit, 2021, p. 279). For example, the discovery of the fast radio bursts in 2007

[14] The "golden event" could itself be subdivided further depending on the individual conditions in which the discovery took place. Examples of completely accidental discoveries can be given, such as those of quasars, pulsars, or gamma ray bursts. Conversely, for some discoveries, there was conjecture regarding their explanation but not yet a complete theory, as in the case of exoplanets. Their existence was not doubted, but the hot Jupiters resulting from the first observations were not yet compatible with established theories. As an intermediate stage between the two extremes, the discovery of the Cosmic Microwave Background can be mentioned, which Gamow had already predicted but which occurred entirely coincidentally and independently of this prediction.

was made with the Parkes radio telescope, which came into operation in 1961. Nonetheless, the existence of a rather strong connection between the introduction of new instruments and the discovery of new phenomena cannot be denied.

Both ideal types of knowledge acquisition are significant for teaching. The theory-based approach (or "logic tradition") corresponds to the classical epistemological view of science. To follow it means to put theory first and deduce conclusions, which must then be tested on empirical data via experiments or planned observations. If teaching is approached in an explorative way, a well-founded overview of the domain of phenomena is necessary to recognize the presence of a relevant phenomenon and its precise characteristics (see Section 2.1.2). Without an appropriate embedding in existing prior knowledge, one otherwise faces the risk of missing any learning effects (Eberbach & Crowley, 2009). Providing the orientational knowledge (Muckenfuß, 2006, pp. 64–66) necessary to enable a systematic exploratory approach is a major challenge in classroom implementation oriented toward this ideal type. The discovery of a "golden event," in contrast, can only be reproduced as the concept cannot be transferred to school contexts in science.[15]

2.2.2 Pure/Experimental

The description of empirical methodological approaches in physics usually includes observation and experiment (Dilling et al., 2020). As we will argue below, the distinction between these two terms is not as straightforward as it may first appear.[16] The very question of when an action can be considered an experiment is by no means uncontroversial. For example, Sandell (2010, p. 262) posits that "astronomers do in fact experiment on the entities they study." According to Sandell, the handling of the observational instrument is decisive for the classification as an experimental action, which is why, for example, the discovery of cosmic background radiation would be considered a consequence of an experiment since "Penzias and Wilson are of a kind with laboratory experimenters" (Sandell, 2010). In the following, we use a different approach, which maintains a separation between observational and experimental science on the one hand and, on the other hand, also shows where the two sides of the methodological approaches converge and mutually determine each other without losing their specific contribution and independent methodological character. In contrast to Sandell, this section will be less concerned with the concrete actions involved in dealing with instruments and take a more abstract view of the question of pure and experimental observations.

Such a separation between observational and experimental actions is appropriate because, from an educational perspective, understanding the need for variable control is an important part of understanding experimental work (Popper, 1976; Schwichow et al., 2016). Therefore, the distinction between the two methods demonstrates the value of experimental work, and fields exist in which the manipulation of the object[17] of study is

[15] On the differences between scientific inquiry in subject science and in the classroom, see, for example, Emden (2021).
[16] Exemplary definitions have already been given for these terms in the introduction.
[17] By a manipulation of the object is meant here that no changes can be made to the state of the source of a signal, in the literal sense of Fig. 4 to be presented below. In this sense, the use of filters in an optical system or the extension to another, visually not directly perceptible, spectral range is not to be seen as a manipulation of the object, since the radiation and transmission up to the filter or instrument is still unperturbed and unperturbable. This distinction is particularly relevant for astronomical examples, where the observed object may already no longer exist when the observer receives and analyzes the incoming light.

simply not possible (e.g., astrophysics and cosmology or geology; Ford, 2005) or not desirable (e.g., biology). The separation of the methods thus also enables a reassessment of their respective value and limits, contributing to the development of a comprehensive understanding of the nature of science (McComas, 1998).

A model that allows such separation of observation and experiment following a clear criterion without creating an unnecessary discontinuity in empirical approaches to nature is described by Harwit (1981). Here, observation is a form of passive information gathering, limited to the data that the system under study provides by itself. However, no sharp boundary is drawn with the experiment since interaction occurs (in the form of volitionally set conditions) but the experimenter continues to passively record only the data provided by the system in the given conditions at a given time. Thus, for Harwit, the observation is a 0th-degree experiment, that is, one in which there is no variable control (Fig. 2).[18] For this characterization, whether the external conditions are deliberately omitted (e.g., observation of the courtship behavior of a wild bird in its natural environment) or such an intervention is in principle not possible (e.g., observation of the effects of the gravitational interaction of two galaxies) is irrelevant.

This model can now be understood in two ways. It could be concluded that observation as an independent method does not exist at all since it is merely a special manifestation of the experiment. Alternatively, observation is an inseparable part of the experiment, which provides conditions that can be systematically controlled but would be deprived of the mandatory feedback to the experimenter without subsequent observation. We assume the second variant. On the one hand, we see an autonomous method realized in the special epistemological properties of observation (see the next section) and, on the other hand, this also corresponds most closely to the common definitions of experiment in the classroom, in which parameter variation is essential (e.g., Boudreaux et al., 2008; Chen & Klahr, 1999).

Another view, which uses a differentiated but fluent account of the terms' relation to each other, is presented by Kunert (1971).[19] In contradiction with Harwit, Kunert considers observation to be a method in its own right. Where variable control is not possible or not desired, he speaks of pure observation. At the other end of the spectrum, we find observation within an experiment, where the term is explicitly not reduced to the experiment alone, but it is emphasized that an experiment is dependent on observation because it is the only way to establish a flow of information back to the experimenter.

In the transitional zone between pure observation and observation within an experiment, Kunert identifies another form of observation. Whenever time appears as a dependent variable in an experiment, the latter can neither be an experiment nor a pure observation. This

[18] Literally, "Observation is the simplest form of experimentation" (Harwit, 1981, p. 17). In another reading, observation could be regarded as the most complex form of experiment, in which access to the object is no longer possible and the experimenter, now to be called observer, is required to have very special skills. The author believes, however, that even under this reading, due to the impossibility of access to the object, a distinction between the methods of experiment and observation can be maintained. Furthermore, the dimensions presented in the further course remain unaffected by the question at which "end" of the experiment the observation follows. The above alternative reading was suggested by one of the anonymous reviewers of this paper, who should be explicitly thanked here for opening this alternative view.

[19] The contribution of this lesser-known author is referred to since it already presented a detailed discussion of the concept of observation at a relatively early time. The work is also recognizably in the tradition of the educational theory of the German Democratic Republic, which in turn entails a certain view of the ways of knowledge in physics teaching, known in the German-speaking world as the "experimental method." However, a detailed discussion of this view cannot be done within the scope of this paper. For this purpose, Kunert differentiates the concept of observation in a way that is not known elsewhere in the literature.

assumption is justified by Kunert's definition of the experiment, in which a "conditional variation in the exploratory arrangement brings about a conditional change" (Kunert, 1971, p. 203). Because time is not assumed to be part of the physical system,[20] there is no opportunity to establish a causal connection between "two property changes of a physical system" (Kunert, 1971, p. 203). Whether the time dependence occurs in explicit (as is usual in kinematics) or implicit form (e.g., in the investigation of processes of thermal conduction) does not matter.

Kunert introduces the necessity of this additional gradation between observation and experiment by means of an example: the observation of the movement of a pendulum clock would constitute a pure observation according to the definition above. Although the pendulum clock is not a natural phenomenon, the oscillation of the pendulum proceeds unaffected when the clock is purely observed. In contrast, an experiment on the pendulum is an experiment according to the common definition. If one follows Kunert, it is difficult to see why so much importance should be attached to the construction and the determination of the conditions and the triggering of the oscillation of the object or the experimental setup in order to deduce the existence of two different kinds of observations. Kunert calls this third form of observation "experimental observation."

A difference remains between the two pendulums shown in Fig. 3 (middle) because Kunert defines experimental observation as.

> a process, for detecting a property change on a physical system, in which the process is initiated, but without inducing a conditional variation determined by the objective of exploration and a change caused thereby. (1971, p. 211, translation by the author)

The intentional triggering of the process is crucial here; however, as Kunert notes, it cannot be sufficient to speak of two separate methods. Implicitly, he is referring to the intention underlying an observation, which Norris (1985) sees as the starting point for approaching the core of an observation.

Here, we agree with Kunert that this example reveals an inconsistency in the dichotomous distinction between observation and experiment. We see in it a confirmation of both the assumption that observation is independent of experiment and the necessity to always conceive observation as a part of any experimental investigation. For Kunert, experimental observation represents a link between pure observation and observation in experiment (Kunert, 1971, p. 204). Abelmann (1965) defends a similar view when she points out methods that merge into each other and the fluid boundaries between them.

To further clarify the difference between pure observation and observation in experiment, it is helpful to return to the historical example of the discovery of X-rays mentioned earlier. The original and incidental discovery was a pure observation: the generation of the glow was not intended within the setup; the screen just happened to be near the tube. Thus, there was no deliberate parameter variation. However, this pure observation was transformed into an experiment without any change to the setup. Here, the character of an experiment is found less in the composition of the material than in the conception and conscious control of the variables. With an observation in an experiment, the phenomenon could now be subject to control of the variables and be tested for reproducibility as well (see Section 2.1.3).

However, the circumstances were completely different when the ongoing question of the possibility of direct perceptibility of the newly discovered radiation directly and alone with

[20] Kunert thus disregards the theory of relativity, which prescribes a coupling of the observer's state of motion to time. Here, it becomes clear that Kunert confines himself entirely to classical mechanics.

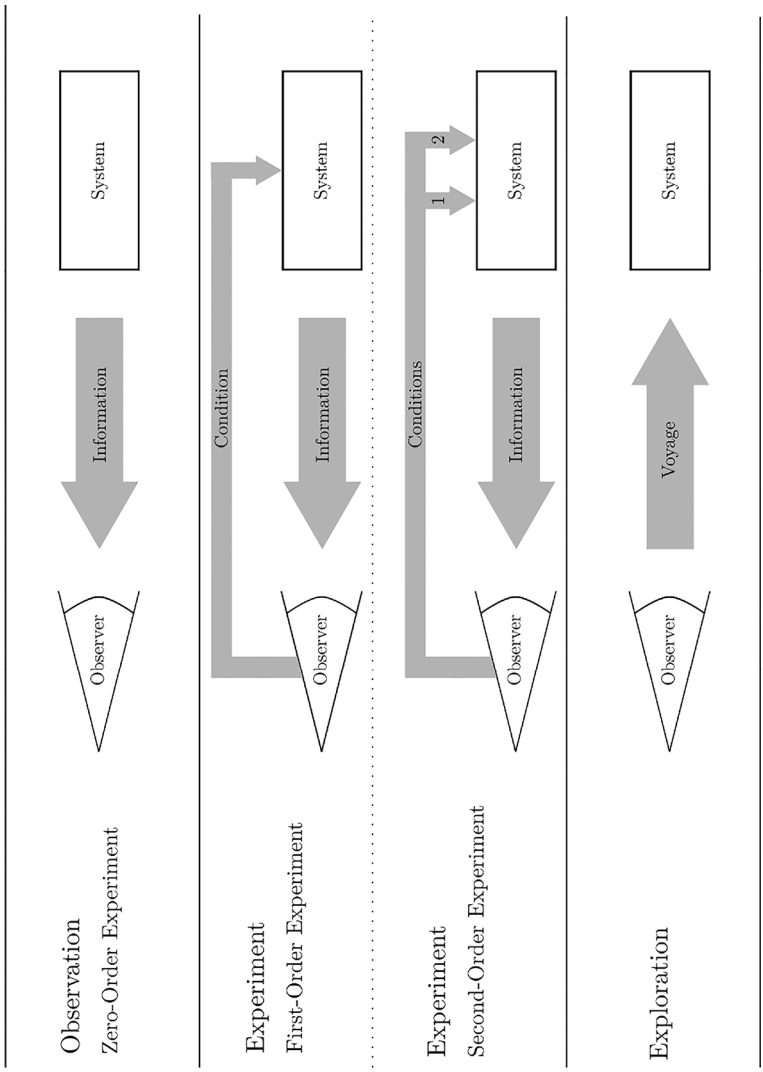


Fig. 2 Relationship between observation and experiment according to Harwit (2021)

Fig. 3 Examples for distinguishing between pure and experimental observation as well as observation in experiment

the eye arose. The complexity of the circumstances (e.g., only partially suitable tubes, the optimal positioning in the room, and the necessary adaptation of the eye to darkness) could not be mastered and controlled at this early stage in the history of this discovery. Because systematic variation was impossible for practical reasons, it is again a pure observation. As such, it is initially subject to much greater fundamental uncertainties as to the causes of the phenomenon than observations in experiments.

From an educational point of view, when distinguishing between pure observations and observation in experiment, creating an awareness of the existence of the different methods of knowledge acquisition seems particularly relevant. Their specific applications and limitations, as well as their close interrelatedness,[21] are essential elements of an understanding of the methodological repertoire and procedures in the sciences.

Hodson (1998) also emphasizes the methodological importance of variable manipulation and the possibility for the experiment to produce phenomena that do not occur in arbitrary conditions. His "passive observation" is congruent with the term "pure observation" used here and is not to be understood in the sense of a passive attitude of the observer:

> Scientific knowledge is at its most powerful and most effective when it is able to control and manipulate events. Indeed, many of the events observed during experiments do not occur in the natural world. In such circumstances, the experimental approach is able to obtain information that is considerably more detailed and precise than that arising from passive observation. (p. 193)

At the same time, however, it is also emphasized that testing theories is possible even in those areas of science where experiments are not feasible or are rejected for ethical reasons (Hodson, 1998; Ioannidou & Erduran, 2021).

2.2.3 Direct/Indirect

The pair of terms direct/indirect refers here to instrument use. Norris (1985) discusses the use of instruments with the terms "simple" and "complex" observations. In addition, Norris raises the question of the extent to which observations are fundamentally tied to the human senses and whether other information channels can be used. With regard to simple observations, he states:

> Scientific observation ranges from relatively simple sorts, such as Darwin's observing the different shapes of the beaks of inches, to relatively complex ones, such as astrophysicists' observing the center of the sun. (Norris, 1985, pp. 831-832)

Furthermore, he contends that teaching based on what he considers simple observations fosters misconceptions about observation as a method, notably the notion that it is necessarily linked to the human senses and is "the simplest mental activity in which scientists engage" (Norris, 1985, p. 832). While we share his apprehension regarding the aspects of the nature of science addressed, we think that the division into simple and complex observations is misplaced, especially in the way it is connected to the cognitive performances involved. For example, the inference of unobservable evolutionary adaptation to external conditions from the observation of certain physical peculiarities of finches on the

[21] This linkage also includes, for example, modeling, which often precedes experimentation and observations for the purpose of theory testing (Emden, 2021).

Galapagos Islands seems extremely complex. In contrast, even a linkage of several measurement instruments can fall short in its cognitive demand, if the mental hurdle for the individual measurement processes is lower overall. Thus, Heath (1980) emphasizes the view of Francis Darwin, who saw the special talent of his father Charles in his ability to perceive what remained hidden to others, "his power of seeing and thinking what most of the world had overlooked" (Rapport & Wright, 1963, as cited in Heath, 1980, p. 156).

Since the pair of terms simple/complex is not very suitable for describing the use of instruments, direct/indirect will be used here instead. This pair also has the advantage of being value neutral, which seems appropriate at the methodological-instrumental level. On the level of teaching practice, however, a classification into simple and complex observation tasks may well be appropriate, in the sense of a ranking of the tasks according to the necessary level of competence or orientation knowledge.

In addition to the individual influences related to the observer, the instruments significantly affect the results of an observation. If observations reach the limits of human perceptual or instrumental capabilities, individual interpretation gains influence. If one understands science as a social construct, the degree of confidence in its results is also dependent on the acceptance of the methods; therefore, the instruments used also play a central role. With the introduction of the telescope into astronomy, Kepler and especially Galileo emphasized its superiority over the eye. Both were interested not merely in providing the eye with an aid but rather in replacing it. For their opponents, perception by the eye was central, which would ultimately be the "final arbiter and measure of all observations" (Gal & Chen-Morris, 2010). To Galileo, however, it was clear that observation without an instrument was by no means to be preferred, even if the possibility existed. For him, such direct observations were "fundamentally suspect" (Gal & Chen-Morris, 2010).

While the introduction of a new instrument is usually met with a certain amount of mistrust, this suspicion changes when the instrument proves itself in everyday scientific practice. The instruments—and, with them, entire processes of data collection and analysis—thus become a black box as they are no longer assigned a particular epistemological role (Pinch, 1985). In educational contexts, however, these developmental steps can rarely be traced, and the functioning of complex instruments is not scrutinized. Consequently, the understanding of the relationship between the instrument and the variable being measured, which is particularly important from an educational perspective, is lost (Götze & Raack, 2022). Additionally, in educational contexts, one often encounters an attitude comparable to Galileo's instrumentalism (Gal & Chen–Morris, 2010) but inappropriate in this context: observations and measurements for which data collection relies more heavily on the human senses (e.g., using an analog multimeter) are usually viewed with skepticism. Meanwhile, digital devices, whose black-box character is even more pronounced,[22] are considered more trustworthy (Götze & Raack, 2022). However, the accuracy of the device is generally not taken into account. The mere fact that an analog scale is present and that the exact measured value may still have to be determined by visual interpolation between two scale lines already casts doubt on the accuracy of the measurement as a whole.

[22] The fact that the underlying process of measurement is significantly more obscure in a digital instrument than in a moving coil instrument is not comprehensible to a completely inexperienced user. Here, only the concrete handling for successful measurement and the type of presentation of the measured value play a role in the evaluation, not the underlying measurement principle. A discussion of these principles can hardly take place within a measurement task.

The reliance on such black-box instruments also leads to alienation from the phenomena themselves. This is particularly problematic in the case of time-varying phenomena (e.g., variable voltage or current waveforms), which are often better identified with analog instruments unless even more complex instruments are used that allow direct conversion to a graphical representation. This is accompanied by a limitation of the possibilities of observation, which is particularly suitable for investigating the processual nature and variability of phenomena as well.

It should be emphasized again that an indirect observation, which uses the various instruments in several intermediate steps, is nevertheless an observation and not simply an experiment if no variable control is performed with respect to the final goal of the observation. An often-cited example (Norris, 1985; Pinch, 1985; Shapere, 1982) is the observation of solar neutrinos. A whole chain of theories is needed for their ultimate detection (Shapere, 1982):

- The theory of nuclear fusion describing the emission of neutrinos
- The theory of the interaction of neutrinos during their flight from the interior of the sun to the detector
- The theory of the detection of neutrinos in the detector

Each of these three steps is a necessary prerequisite for the successful observation of neutrinos, and each of these theories is complex.[23] Thus, even the detection of the reaction products in the detector fluid involves many assumptions and the use of complex measurement technology. Neutrinos were eventually detected via the reaction of a Geiger counter, which was able to detect the radioactive argon-37 isotope produced by the interaction of the neutrino and the detector liquid (Fig. 4).

For our argumentation, it is important to note that at the end of a long chain of investigations, assumptions, and interpretations lies evidence of something that is clearly not directly observable in itself: the nuclear reactions inside a star. In addition, human senses play a subordinate role at best within this chain. Nevertheless, the scientists involved speak of an observation (Norris, 1985).

2.2.4 Qualitative/Quantitative

The question of the classification of a method of inquiry as qualitative or quantitative is directly that of the use of instruments and the distinction between pure observations and observations in experiments: can there be a pure and at the same time quantitative observation? Or, to put it differently, does the performance of a measurement preclude speaking of an observation?

This question was addressed by Kunert, who first approached the concept of measurement[24] and proposed a broad and a narrow definition:

[23] The seemingly least complex intermediate step—the description of the interaction during the information transfer, which assumed that no interaction occurs and the neutrinos arrive at the detector unchanged in number and properties—turned out to be wrong.

[24] The concept of measurement according to Kunert must be supplemented by other aspects with regard to quantum physics. For example, measurement is no longer possible independently of the state of the system since a feedback on the observed system is unavoidable. Therefore, the extent to which the notion of observation as used in this paper should be applied to quantum physics becomes a question. It seems advisable to speak of measurements, emphasizing their invasive character, and avoid the term "observation" here. Kunert, along with the theory of relativity, does not take this into account. In this contribution, however, the problem touched upon here cannot be treated exhaustively.

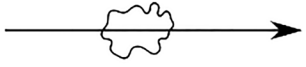
Source	Path	Reciever
theory of the source	theory of the interaction	theory of the reciever
nuclear reaction in the interior of the sun	no interaction during information transfer	interaction with the detectors fluid
$^{1}H^{+} + {}^{1}H^{+} \rightarrow {}^{2}H^{+} + e^{+} + \nu_e$		$\nu_e + {}^{37}Cl \rightarrow {}^{37}Ar + e^{-}$

Fig. 4 Theoretical assumptions in the detection chain for solar neutrinos as an example of indirect observation (Shapere, 1982)

> Measurement in a broader sense: determination of the extensity of a property change, i.e., whether the property change is present or not (qualitative mode of observation: mapping on a dichotomous nominal scale).
> Measurement in the narrower sense: determination of the intensity of a property change, i.e., to what degree the change is present (quantitative method of observation: Mapping on an ordinal or metric scale). (1971, p. 187, translation by the author)

In a measurement in the broader sense, according to Kunert, the senses of the observer often fulfill the role of the measuring instrument.[25] Furthermore, any form of observation (pure and experimental observations and observations in experiment) can be carried out qualitatively and quantitatively (Fig. 5). This argues for seeing in measurements a mere extension and closer specification of the underlying methodological action and, thus, of the method.

Popper expresses a critical view of the role of measurement in the verification of theories. Although he underlines the high value of quantitative statements for such purposes, he also stresses the following:

> We thus explain the role of measurement as an aid to the verification of theories; an aid which, while becoming increasingly important in the course of scientific progress, must never be used to characterize science or theorizing in general because of its late and conditional appearance, and because it is always dependent on theoretical presuppositions. (1964, p. 99, translation by the author)

Here, too, the method of measurement is not equal to the other methods of knowledge but supplements and specifies them.

Thus, it becomes clear that whether a quantitative measurement was performed cannot be relevant for the characterization of the method.[26] Even a pure and quantitative observation (i.e., a measurement without the possibility of variable control) is an observation. Looking at the experiment, the argumentation for this characterization becomes clearer: to classify an action as experimental, any additional measuring activity is irrelevant. The question of whether a quantitative measurement (i.e., a measurement in the narrower sense according to Kunert) occurs during an experiment cannot prevail over the question of variable control or reproducibility. Otherwise, a multitude of experiments would no longer be regarded as experiments. This would affect a large number of experiments whose purpose is to evoke and visualize a phenomenon or that are used in the introductory phase of teaching for motivational reasons. If observation is recognized as a form of knowledge acquisition of equal rank to experimentation,[27] it cannot be subsumed under measurement or equated with it, either generally or in certain situations only.

An example illustrating the smooth transition from qualitative to quantitative observations is provided by astronomy. The foundation for orientation in the starry sky is the

[25] From Norris's (1985) perspective, the strong emphasis on the human senses is a naïve notion of observation. However, it should be noted that Kunert takes a practical view of teaching that also refers to the 1960s and early 1970s. At the time, this assessment likely corresponded to classroom reality. Furthermore, other forms of information registration are not excluded in principle, but rather it is emphasized that with technical development, methods of automatic registration will also gain importance in the lessons (Kunert, 1971, p. 188).

[26] Malik (2017), for example, also provides a number of examples of observations, most of which are measurements (also explicitly named as such). These are, moreover, predominantly quantitative in nature.

[27] So far, we have not been able to present any arguments against such an assumption. However, a systematic comparison in structured contrast falls outside of the scope of this paper.

constellations or asterisms as the imaginary connecting lines between bright stars.[28] Finding such distinctive arrangements of stars is a basic observation task in astronomy courses. Cassiopeia is a constellation of the northern sky, which is quite conspicuous with its prominent "W" shape. Observations over a night or a whole year show the change of its orientation in the sky (another classical observation task). One can hardly avoid noticing that Cassiopeia consists of five comparably bright stars that form the well-known figure. Is this observation already a measurement, in the form of a number of stars related to a certain figure?

A more advanced task of an astronomy course could be to determine the number of visible stars in the sky. This task could be approached by counting the number of visible stars in the vicinity of the prominent constellations according to their modern definition. Regardless of the practical difficulties, the indication of the result would already be a kind of number density representing stars per constellation. Such a task could be extended by introducing a "star counting cone" (Winnenburg, 1996) through which a defined fraction of the total area of the sky is visible, in which the stars are counted.[29] The repeated counting and averaging in different areas of the sky thus results in a real number density of the visible stars and allows extrapolation to the whole celestial sphere.[30] Without any doubt, the last step is a measurement in the form of a number density with the unit $\frac{stars}{deg^2}$. However, a sharp, previously made subdivision into observation and measurement does not seem to be expedient since a clearly definable limit does not exist (Fig. 6). If, however, one speaks of an increasing quantification of the observation, no classification problem arises.[31]

2.2.5 Iconic/Numerical

The question of how the chosen form of representation of the observed object (e.g., in analogy experiments) or—in the case of indirect observations—of the data influences their reception and interpretation by the observer cannot be treated exhaustively in this paper. Therefore, we limit ourselves to outlining the field and pointing out some basic aspects.

The conceptual pair iconic/numeric occupies a position in transitional zones in two respects. First, as concerns the transition from the methodological-instrumental to the methodological-instructional level, the design of the lesson (i.e., the forms of representation used) is generally incumbent on the teacher and, thus, subject to their conscious decision to resort to one or another form. This decision (e.g., to represent measured values on analog or digital measuring devices) is therefore a practical teaching one and is largely independent of the phenomena.

[28] These asterisms have nothing to do with the modern definition of constellations, which are defined by lineaments between celestial coordinates, enclosing the classical constellations but, in a sense, drawing a frame around the conspicuous structures. We use the term "constellation" here as a synonym for asterism.

[29] A method that is similar in principle was used by William Herschel (1817) to determine the structure of the Milky Way.

[30] When extrapolating, the fact that the stars are not evenly distributed and the main plane of the Milky Way may cause distortions must be taken into account.

[31] It could be argued that a "real measurement" occurs exactly at the moment when a mathematically exactly defined reference area is taken as a basis. However, since there is no change in the fundamental action—stars are still counted, and this counting concerns a more or less exactly defined section of the sky—we do not recognize any fundamental change in this action. In contrast, it seems appropriate to speak of a quantitative observation or a measurement in the narrower sense according to Kunert (1971) from this moment on.

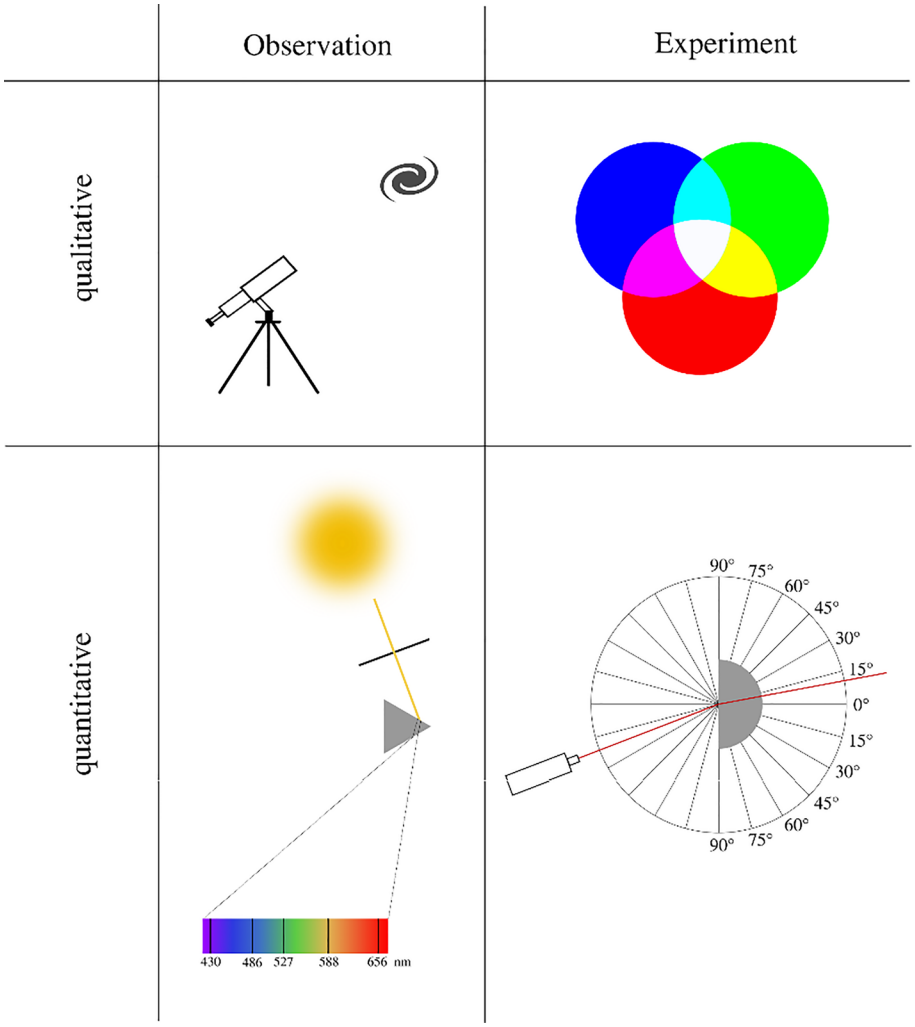


Fig. 5 Examples of qualitative and quantitative observations and experiments

The second transitional area is that between qualitative and quantitative observation. The choice of the form of representation is closely related to the intention to represent a measured value in a certain way and primarily concerns the subarea of quantitative observations (i.e., measurements). As we have explained, we regard the transition from qualitative to quantitative observations as fluid (see Section 2.2.4). However, the question of the form of representation should rarely arise in a pure qualitative observation if the phenomenon itself is observed rather than a representation of it. However, if one also understands observation in an experiment as a form of observation, as we have done here, how to present measured values should also be included in the educational reflection on the method of observation.

There is a close relationship to the question of the instrument and, thus, to the conceptual pair direct/indirect. Like in science (Pinch, 1985), the measurement instruments used have a

Fig. 6 Transition from qualitative to quantitative observation

black-box character for the learners in the classroom (Gravemeijer et al., 2017, p. 9). However, in line with what has been presented so far, it seems desirable to reveal how instruments work, for example, by looking at the history of science (see Section 2.2.3). As we stated at the beginning, observation is particularly suitable for investigating processes (i.e., time-varying phenomena), and it is precisely here that the chosen form of representation has a particularly important role to play (e.g., in the introductory experiment on the induction of a voltage in a coil under the influence of a variable magnetic field). The instruments partly create the conditions in which the observed phenomena occur (Muckenfuß, 2000, 2006, pp. 337–338). Consequently, the choice of a suitable instrument is crucial and also includes that of the form of representation. Götze and Raack show the educational advantages of using analog rather than digital measuring instruments to study electric current:

1. The scale of an analog multimeter is similar to a length measurement that students become familiar with from an early age.
2. The zero point in the scale center enables conclusions about current direction.
3. The velocity of the indicator movement shows the rate of change.
4. The functioning of the measuring instrument can be explained with magnetism.
5. Fluctuating readings on the digital display confuse the students (2022, p. 82).

Points 3 and 5 are particularly relevant to observation.[32] For example, rates of change can hardly be displayed by a simple digital device; as a result, the phenomenon is not visible and eludes observation. Additionally, the fluctuation of the display can distract the untrained student from the essential correlations between the quantities, that is, direct their attention to the influence of the instrument, which rarely deserves such attention. At the same time, the increased use of digital instruments creates a level of confidence in the measurements that is neither justified by the accuracy of the instrument nor by the experimental design itself (Götze & Raack, 2022). However, this assessment of the instrument seems to be predominantly due to the form of representation alone rather than the

[32] In a broader sense, all questions are relevant since the aforementioned advantages are found in a subset of the measurement process that concerns the subject's interaction with the instrument and, thus, can also be understood as an observation of the instrument in a broader sense. Here, we limit ourselves to those subaspects that are more closely related to access to the underlying phenomenon.

functioning of the instrument. The latter appears to be a black box, and the flow of information during the observation is thus essentially shaped by the form of representation.

2.2.6 Summary: Dimensions of Observation at the Methodological-Instrumental Level

As we have shown, observation can be used in a broad methodological spectrum. Like experimental methods, it can be applied in explorative and in validating forms. This constitutes a further indication that the two methods should be regarded as equivalent and observation should not be understood as a preliminary methodological stage in an experimental investigation. Additionally, a wide array of objects of observation exists. Except for pure observations, for which a setting and the modification of conditions are not possible or desired, observations always take place in the context of experiments or in the intermediate form of experimental observation. Thus, observation permeates every method of empirical knowledge acquisition in physics. Without it, access to nature remains blocked. In addition, observations can be made using instruments or without any instrumental support, and the evaluation can thus be qualitative or quantitative. However, even creating chains of instruments does not change the fundamental character of the method (e.g., for a pure observation, since the object of the investigation remains inaccessible). Moreover, the question of whether measurements are made is ultimately secondary and can be used to propose a closer characterization but does not concern the core peculiarities of the method.

Other aspects of the use of observation as a method of knowledge acquisition are often subject to the teacher's decisions when designing the lesson. Individual characteristics of the method can be specifically emphasized. For this, however, the teacher must be aware of the existence of the various dimensions to convey an accurate picture of the method as a whole and be able to explain its possibilities and limitations in detail.

3 Conclusions

Even simple attempts to define the term "observation" show that it is a multi-layered scientific method and differs greatly in its characteristics depending on the situation. The presentation of its characteristics by means of pairs of terms has turned out to be a practicable way to approach a description without claiming completeness. These pairs of terms hereby represent different dimensions of observation. For further structuring, these pairs were divided into two levels, the fundamental and the methodological-instrumental. This was done to separate dimensions that are essential for the method from those that can be used for a more detailed characterization. On the fundamental level, the pairs of terms are opposites, so that they can only be one-sided if an action is to be considered an observation in the scientific sense. On the methodological-instrumental level, in contrast, the pairs represent the two ends of a continuum and thus demonstrate the wide range of possible empirical practices within the sciences. The different orientation towards one or the other end within the dimension may be typical for the respective discipline (physics or biology) or sub-discipline (physics or astrophysics) and thus reflect the character of the respective subject. Regarding teaching, there might also be differences in the practical implementation of a single topic or the handling of certain equipment (as explained in the section on the pair iconic/numeric).

The two-level division made here can be placed in relation to the conceptualizations of nature of science that currently dominate the discourse. These two views are referred to here, according to Kampourakis (2016), as the "general aspects" and "family resemblance" conceptualizations. The general aspects conceptualization represents the consensus among science educators and is based on lists of general aspects of nature of science. Its core ideas address misconceptions that students have about the nature of the sciences. The family resemblance approach, in contrast, aims primarily to express the heterogeneity of the sciences and to counteract the spread of stereotypical views through a truncated representation of NOS aspects in the classroom (Kampourakis, 2016). In addition, the idea of science as a social enterprise, with its own rules and norms, is strongly emphasized (Irzik & Nola, 2022).

The levels used here could be understood as a mapping to these NOS levels, even though the wording of the terms has not been included in the two conceptualizations so far. The fundamental level summarizes minimal requirements for observations to be considered a scientific observation. The one-sided assignment of the pairs of terms here corresponds to a consensus in the literature about the character of the method. Accordingly, on this level, this characterization could be assigned to the general aspects conceptualization.

In contrast, the methodological-instrumental level, with the dimensions to be understood here as a continuum, shows a very heterogeneous picture of observation. This argues for an assignment of this level to the family resemblance approach. In line with this, Irzik and Nola (2022) also refer to the complexity of experiments and observations:

> One can then point out that neither observation nor experiment is a unitary concept by drawing attention to the differences among scientific observations and among scientific experiments, thanks to the new application of the idea of family resemblance to them.

In the present paper, the social aspects, which are strongly emphasized in the family resemblance conceptualization, have remained largely unconsidered.[33] An example is the use of third-party observation data and the associated question of the trustworthiness of the original observers. Multiple historical examples can be cited for this aspect, which can be used to demonstrate the social constitution of science. Equally relevant is the inverse influence of expectations on observations generated by others' reports. A well-known historical example in this regard is the supposedly successful observation of the Martian canals, which was originally caused by the faulty translation of Schiaparelli's observational report and subsequently solidified by the widely claimed confirmation of their existence. The educational problems, which can be caused by such a steering of the attention, need not be emphasized separately here.

At the fundamental level, conscious perception of a phenomenon and mental activity are necessary conditions for observation. In addition, it was shown that also their repetition or at least the reproducibility is necessary to equip the method of observation with a quality criterion which is comparable to the experiment. This is the only way to counteract von Guericke's strong emphasis on the experiment for the process of knowledge. The question of repeatability already partly leads away from the fundamental level, since different forms of observation are already addressed here. Examples of such observations,

[33] Due to the high complexity of personal relationships and social structures, representation in terms of pairs of terms seems less suitable for this purpose. Moreover, the topic forms a complex of its own, which cannot be discussed within the scope of this article.

which like experiments can be considered reproducible, have been presented in this paper as experimental observations.

In astrophysics, in contrast, where mostly only pure observations are possible, certainty about an observed phenomenon does not arise from repeated single observations by different observers or with different instruments. Instead, astrophysical theories gain their increasing certainty from observations of large populations of similar objects, which serve to confirm the predictions many times over. Here too, however, the results of the method can thus be regarded as sufficiently reliable to be in no way inferior to an experimental science.

Not to be confused with the claim to objectivity is the term pair biased/unbiased. For this pair, it was shown that observations are necessarily subject to bias, without this contradicting the claims of science as a cognitive-epistemic system (Irzik & Nola, 2022), emphasizing as essential aims, among others, precisely truth and consistency. Rather, biases must be understood, made conscious, and accounted for as a given component of the observations in the situation, to reach this consistency and truth.

The pair of terms on the methodological-instrumental level show the wide spectrum and the heterogeneity within the observational sciences and indicate in addition, where linking points to the experimental approaches are present. Thus, the example of the discovery of X-rays makes it clear that the forms of observation (pure and experimental) are not only separated by subtle epistemological differences but also have demonstrable effects on the practical work of the scientist in corresponding situations: the scientist himself does not trust his singular observations if a deeper epistemological reflection does not arise. Therefore, the question arises of whether and in which circumstances even pure observation—as the ideal form of observation—can represent a reliable source of knowledge.

The heterogeneous approaches span, for example, exploratory and theory-driven approaches as well as pure and experimental observations and their respective intermediate stages. In the context of the pair direct/indirect, Irzik and Nola (2022) also point out the significant fact that observations do not always require an observer but can also be performed by automated data collection. Again, however, there is a continuum that leads us from unaided observations, through the use of increasingly complex tools, eventually to such completely indirect observations. Nevertheless, these are still actions that have to be considered as observations.

The pairs of terms qualitative/quantitative and iconic/numeric are of particular relevance for schools, as they touch on central actions or modes of representation in the classroom. The relationship between the concept of measurement and observation has been shown here, and the complexity of the concept of measurement itself has been briefly touched upon. Somewhat apart from the NOS conceptualizations is the pair of terms iconic/numeric, which seems to be of increasing importance also in light of the rapid change in modes of representation due to digitization in measurement acquisition and evaluation. Previous approaches to the effects originate predominantly from the field of mathematics education. However, as briefly touched upon, the question of representation seems to be of great importance for the sciences as well.

In this paper, we showed that observation is not a simple way of accessing nature compared to other methods of knowledge acquisition. This finding is in line with earlier works (e.g., Norris, 1985), in which the cognitive demand of the method was already pointed out. In addition, there is a broad spectrum of applications of observation, which also enables the study of phenomena that are not accessible to classical, experimental methods.

The objective of this paper was to show the possibilities and limitations of the method of observation from a fundamental and context-specific point of view. In school as well as in teacher training, a characterization of the method should be provided to present its

complexity and counteract misconceptions of a simple, cognitively undemanding method that is of little benefit for the progress of knowledge.

To know the methodological repertoire of one's subject in detail and internalize the demarcations and transitions between the methods must be an essential goal of teacher training. Only in this way can a necessary level of education be achieved, notably with regard to the nature of science. For the method of observation, this includes knowledge of the possible applications (explorative and validating), the manifestations (pure and experimental observations, observations in experiments), and references to the concept of measurement. Possible interactions with the technical design of instruments and different forms of presentation are a necessary part of sound teacher training.

Acknowledgements The author gratefully acknowledges Mr. Oliver Schwarz for his very profitable feedback on an earlier version of the manuscript. The author also would like to thank the three anonymous reviewers of this journal for their detailed and thoughtful feedback on the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability A corresponding explanation is not applicable to this article.

Declarations

Conflict of Interest The author declares no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abelmann, X. (1965). Zum Verhältnis von Beobachtung und Experiment [On the relationship between observation and experiment]. *Rostocker Philosophische Manuskripte, 2*, 179–186.

Auger, P., Ehrenfest, P., Maze, R., Daudin, J., & Fréon, R. A. (1939). Extensive Cosmic-Ray Showers. *Review of Modern Physics, 11*(3–4), 288–291. https://doi.org/10.1103/RevModPhys.11.288

Boudreaux, A., Shaffer, P. S., Heron, P. R. L., & McDermott, L. C. (2008). Student understanding of control of variables: Deciding whether or not a variable influences the behavior of a system. *American Journal of Physics, 76*(2), 163–170. https://doi.org/10.1119/1.2805235

Brown, R. A., & Kumar, A. (2013). The scientific method: Reality or myth? *Journal of College Science Teaching, 42*(4), 10–11.

Burian, R. M. (1997). Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1938–1952. *History and Philosophy of the Life Sciences*, *19*(1), 27–45. Retrieved from http://www.jstor.org/stable/23332033

Burian, R. M. (2013). Exploratory experimentation. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 720–723). Springer New York. https://doi.org/10.1007/978-1-4419-9863-7_60

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120.

Chi, M. T. H., Hutchinson, J. E., & Robin, A. F. (1989). How inferences about novel domain–related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly, 35*(1), 27–62.

Daston, L. (2001). *Eine kurze Geschichte der wissenschaftlichen Aufmerksamkeit [A brief history of scientific attention]. Themen: Vol. 71*. Carl Friedrich von Siemens Stiftung.

Daston, L., & Lunbeck, E. (2011). Introduction: Observation observed. In L. Daston & E. Lunbeck (Eds.), *Histories of scientific observation* (pp. 1–9). The University of Chicago Press.

Dilling, F., Stricker, I., Tran, N. C., & Vu, D. P. (2020). Development of knowledge in mathematics and physics education. In S. F. Kraus & E. Krause (Eds.), *MINTUS – Beiträge zur mathematisch–naturwissenschaftlichen Bildung. Comparison of Mathematics and Physics Education I: Theoretical Foundations for Interdisciplinary Collaboration* (1st ed., pp. 299–344). Springer Spektrum.

Duncombe, R. L. (1945). Personal equation in astronomy. *Popular Astronomy*, *53*(1), 2–13, 63–76, 110–120.

Eberbach, C., & Crowley, K. (2009). From everyday to scientific observation: How children learn to observe the biologist's world. *Review of Educational Research, 79*(1), 39–68. https://doi.org/10.3102/0034654308325899

Einstein, A. (1955). *Mein Weltbild. Ullstein Bücher: Nr. 65*. Frankfurt/M.: Verl. Das goldene Vlies.

Emden, M. (2021). Reintroducing "the" scientific method to introduce scientific inquiry in schools? *Science & Education, 30*(5), 1037–1073. https://doi.org/10.1007/s11191-021-00235-w

Ford, D. J. (2005). The challenges of observing geologically: Third graders' descriptions of rock and mineral properties. *Science Education, 89*(2), 276–295. https://doi.org/10.1002/sce.20049

Furtak, E. M., & Penuel, W. R. (2019). Coming to terms: Addressing the persistence of "hands–on" and other reform terminology in the era of science as practice. *Science Education, 103*(1), 167–186. https://doi.org/10.1002/sce.21488

Gal, O., & Chen-Morris, R. (2010). Empiricism without the senses: How the instrument replaced the eye. In C. T. Wolfe & O. Gal (Eds.), *The Body as Object and Instrument of Knowledge* (Vol. 25, pp. 121–147). Springer, Netherlands. https://doi.org/10.1007/978-90-481-3686-5_7

Galison, P. (1997). *Image and Logic*. University of Chicago Press.

Gillies, D. (2006). Kuhn on discovery and the case of penicillin. In W. J. González & J. Alcolea (Eds.), *Contemporary perspectives in philosophy and methodology of science* (pp. 47–63). Netbiblo.

Glasser, O., Röntgen, W. C., & Boveri, M. (1933). *Wilhelm Conrad Röntgen and the early history of the Roentgen Rays: With a chapter: Personal reminiscences of W. C. Röntgen*. Bale & Danielsson.

Götze, D., & Raack, P. (2022). Comparison: Numbers, quantities and units. In F. Dilling & S. F. Kraus (Eds.), *Comparison of Mathematics and Physics Education II* (pp. 79–90). Springer Spektrum. https://doi.org/10.1007/978-3-658-36415-1_5

Gravemeijer, K., Stephan, M., Julie, C., Lin, F.-L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education, 15*(S1), 105–123. https://doi.org/10.1007/s10763-017-9814-6

Grinnell, F. (2013). Research integrity and everyday practice of science. *Science and Engineering Ethics, 19*(3), 685–701. https://doi.org/10.1007/s11948-012-9376-5

Harwit, M. (1981). *Cosmic discovery: The search, scope, and heritage of astronomy*. Harvester Press.

Harwit, M. (2021). *Cosmic messengers: The limits of astronomy in an unruly universe*. Cambridge University Press.

Haslam, F., & Gunstone, R. (Eds.) (1998). *The influence of teachers on student observation in science classes* [Paper presentation]. Annual Meeting of the National Association for Research in Science Teaching. San Diego.

Hazard, C., Jauncey, D., Goss, W. M., & Herald, D. (2014). The sequence of events that led to the 1963 publications in Nature of 3C273, the first quasar and the first extragalactic radio jet. *Proceedings of the International Astronomical Union, 10*(S313), 183–189. https://doi.org/10.1017/S1743921315002173

Heath, T. (1980). Observation, perception and education. *European Journal of Science Education, 2*(2), 155–160. https://doi.org/10.1080/0140528800020206

Herrmann, D. B. (2004). *Astronomiegeschichte: Ausgewählte Beiträge zur Entwicklung der Himmelskunde [History of Astronomy: Selected Contributions to the Development of Celestial Science]*. PAETEC.

Herschel, W. (1817). XXIV. Astronomical observations and experiments tending to investigate the local arrangement of the celestial bodies in space, and to determine the extent and condition of the Milky Way. *Philosophical Transactions of the Royal Society of London, 107*, 302–331. https://doi.org/10.1098/rstl.1817.0025

Hidi, S. (2001). Interest, Psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 7712–7715). Pergamon. https://doi.org/10.1016/B0-08-043076-7/01535-7

Hodson, D. (1998). Science fiction: The continuing misrepresentation of science in the school curriculum. *Curriculum Studies, 6*(2), 191–216. https://doi.org/10.1080/14681369800200033

Holton, G. J. (1979). Einsteins Methoden zur Theoriebildung [Einstein's Methods for Theory Formation]. In P. C. Aichelburg, R. U. Sexl, & P. G. Bergmann (Eds.), *Albert Einstein: Sein Einfluss auf Physik, Philosophie und Politik* (pp. 111–140). Vieweg.

Ioannidou, O., & Erduran, S. (2021). Beyond hypothesis testing. *Science & Education, 30*(2), 345–364. https://doi.org/10.1007/s11191-020-00185-9

Irzik, G., & Nola, R. (2022). Revisiting the foundations of the family resemblance approach to nature of science: Some new ideas. Science & Education. https://doi.org/10.1007/s11191-022-00375-7

Johnson, K. E., & Mervis, C. B. (1994). Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. *Developmental Psychology, 30*(3), 418–435. https://doi.org/10.1037/0012-1649.30.3.418

Johnson, K. E., & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*(3), 248–277. https://doi.org/10.1037/0096-3445.126.3.248

Kampourakis, K. (2016). The "general aspects" conceptualization as a pragmatic and effective means to introducing students to nature of science. *Journal of Research in Science Teaching, 53*(5), 667–682. https://doi.org/10.1002/tea.21305

Kim, S., & Park, J. (2018). Development and application of learning materials to help students understand ten statements describing the nature of scientific observation. *International Journal of Science and Mathematics Education, 16*(5), 857–876. https://doi.org/10.1007/s10763-017-9823-5

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211–228. https://doi.org/10.1037/0033-295X.94.2.211

Knogler, M. (2017). Situational interest: A proposal to enhance conceptual clarity. In P. A. O'Keefe & J. M. Harackiewicz (Eds.), *The Science of Interest* (pp. 109–124). Springer International Publishing. https://doi.org/10.1007/978-3-319-55509-6_6

Kowal, C. T., & Drake, S. (1980). Galileo's observations of Neptune. *Nature, 287*(5780), 311–313. https://doi.org/10.1038/287311a0

Kraus, S. F. (2018). Die Beobachtung im Gefüge des Erkenntnisprozesses der Physik [The observation in the structure of the process of knowledge acquisition of physics]. In V. Nordmeier & H. Grötzebauch (Eds.), *PhyDid B, Didactics of physics, contributions to the DPG Spring Conference Dresden, 2018.*

Kuhn, T. S. (1962). *The structure of scientific revolutions*. *International encyclopedia of unified science (Vol. 2), Foundations of the unity of science: No. 2*. Chicago University Press.

Kunert, E. (1971). *Zur Praxis und Theorie der empirischen Verfahren im Physikunterricht der allgemeinbildenden polytechnischen Oberschule* [On the practice and theory of empirical methods in the teaching of physics in general polytechnic high school.]*: – ein Beitrag zur Schaffung von Grundlagen für de Weiterentwicklung der Theorie des physikalischen Lehrgeräts*. Dissertation. Pädagogische Hochschule Potsdam.

Malik, S. (2017). Observation versus experiment: An adequate framework for analysing scientific experimentation? *Journal for General Philosophy of Science, 48*(1), 71–95. https://doi.org/10.1007/s10838-016-9335-y

McComas, W. F. (1998). The principal elements of the nature of science: Dispelling the myths. In W. F. McComas (Ed.), *Science & Technology Education Library. The Nature of Science in Science Education.* Kluwer Academic Publishers.

Muckenfuß, H. (2000). Retten uns die Phänomene? Anmerkungen zum Verhältnis von Wahrnehmung und Theorie [Do phenomena save us? Notes on the Relationship between Perception and Theory.]. *Plus LUCIS, 3*, 10–14.

Muckenfuß, H. (2006). *Lernen im sinnstiftenden Kontext: Entwurf einer zeitgemäßen Didaktik des Physikunterrichts [Learning in a Meaningful Context: Design of a Contemporary Didactics of Physics Education]*. Cornelsen.

Nakajima, M., Schmitt, L. I., & Halassa, M. M. (2019). Prefrontal cortex regulates sensory filtering through a basal ganglia–to–thalamus pathway. *Neuron, 103*(3), 445-458.e10. https://doi.org/10.1016/j.neuron.2019.05.026

Norris, S. P. (1985). The philosophical basis of observation in science and science education. *Journal of Research in Science Teaching, 22*(9), 817–833. https://doi.org/10.1002/tea.3660220905

O'Malley, M. A. (2007). Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case. *History and Philosophy of the Life Sciences, 29*(3), 337–360.

Persson, L. (1996). Pierre Auger–A life in the service of science. *Acta Oncologica, 35*(7), 785–787. https://doi.org/10.3109/02841869609104027

Pinch, T. (1985). Towards an analysis of scientific observation: The externality and evidential significance of observational reports in physics. *Social Studies of Science, 15*(1), 3–36. https://doi.org/10.1177/030631285015001001

Popper, K. R. (1964). Naturgesetze und theoretische Systeme [Laws of nature and theoretical systems]. In H. Albert (Ed.), *Theorie und Realität* (pp. 87–102). Mohr.

Popper, K. R. (1976). *Logik der Forschung* [Logic of research]. *Die Einheit der Gesellschaftswissenschaften: Vol. 4*. Mohr.

Röntgen, W. C. (1897). Weitere Beobachtungen über die Eigenschaften der X Strahlen (Nouvelles observations sur les propriétés des rayons X). *Sitzungsberichte der Akademie der Wissenschaften zu Berlin*, 576–592.

Rapport, S., & Wright, H. (1963). *Science: Method and meaning*. New York University Press.

Sandell, M. (2010). Astronomy and experimentation. *Techné: Research in Philosophy and Technology*, *14*(3), 252–269. https://doi.org/10.5840/techne201014325

Schaffer, S. (1988). Astronomers mark time: Discipline and the personal equation. *Science in Context, 1*, 115–145.

Schmidt, M. (1963). 3C 273: A star-like object with large red-shift. *Nature, 197*(4872), 1040. https://doi.org/10.1038/1971040a0

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta–analysis. *Developmental Review, 39*, 37–63. https://doi.org/10.1016/j.dr.2015.12.001

Selby, C. C. (2006). What makes it science? A modern look at scientific inquiry. *Journal of College Science Teaching, 35*(7), 8–11.

Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science, 49*(4), 485–525.

Simonyi, K. (2012). *Kulturgeschichte der Physik: Von den Anfängen bis heute [Cultural History of Physics: From the Beginnings to Today]*. Deutsch.

Slaby, A. K. H. (1907). *Festvortrag aus Anlass der Grundsteinlegung des Deutschen Museums zu München gehalten im Wittelsbach–Palais am 13. November 1906* [Lecture given on the occasion of the groundbreaking ceremony of the Deutsches Museum in Munich at the Wittelsbach-Palais on November 13, 1906.]. https://doi.org/10.1007/978-3-662-32761-6

Steinle, F. (1997). Entering New Fields: Exploratory Uses of Experimentation. *Philosophy of Science, 64*, 65–74.

Thomson, J. J. (1894). On the velocity of the cathode–rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 38*(233), 358–365. https://doi.org/10.1080/14786449408620644

Tomkins, S. P., & Tunnicliffe, S. D. (2001). Looking for ideas: Observation, interpretation and hypothesis–making by 12–year–old pupils undertaking science investigations. *International Journal of Science Education, 23*(8), 791–813. https://doi.org/10.1080/09500690010016049

Torretti, R. (1986). Observation. *British Journal for the Philosophy of Science, 37*, 1–23.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*(1), 123–183. https://doi.org/10.1016/0364-0213(94)90022-1

Wachtel, P. L. (1967). Conceptions of broad and narrow attention. *Psychological Bulletin, 68*(6), 417–429. https://doi.org/10.1037/h0025186

White, D. W. (2022). Building an astrophysics/astrochemistry laboratory from scratch. *The Physics Teacher, 60*(5), 362–364. https://doi.org/10.1119/10.0010394

Winnenburg, W. (1996). Die Sternzählbox [The star counting box]. *Astronomie + Raumfahrt Im Unterricht*, *33*(1), 18.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.